J Biol Inorg Chem (2014) 19:903–912
DOI 10.1007/s00775-014-1125-6

ORIGINAL PAPER

Solution structure and metal ion binding sites of the human CPEB3 ribozyme's P4 domain

Miriam Skilandat · Magdalena Rowinska-Zyrek · Roland K. O. Sigel

Received: 8 February 2014 / Accepted: 5 March 2014 / Published online: 21 March 2014

Abstract Three ribozymes are known to occur in humans, the CPEB3 ribozyme, the CoTC ribozyme, and the hammerhead ribozyme. Here, we present the NMR solution structure of a well-conserved motif within the CPEB3 ribozyme, the P4 domain. In addition, we discuss the binding sites and impact of Mg^{2+} and $[Co(NH_3)_6]^{3+}$, a spectroscopic probe for $[Mg(H_2O)_6]^{2+}$, on the structure. The well-defined P4 region is a hairpin closed with a UGGU tetraloop that shows a distinct electrostatic surface potential and a characteristic, strongly curved backbone trajectory. The P4 hairpin contains two specific Mg^{2+} binding sites: one outer-sphere binding site close to the proposed CPEB3 ribozyme active site with potential relevance for maintaining a compact fold of the ribozyme core, and one inner-sphere binding site, probably stabilizing the tetraloop structure. The structure of the tetraloop resembles an RNase III recognition structure, as previously described for an AGUU tetraloop. The detailed knowledge of the P4 domain and its metal ion binding preferences thus brings us closer to understanding the importance of Mg^{2+} binding for the CPEB3 ribozyme's fold and function in the cell.

Keywords NMR · CPEB3 · Ribozyme · Mg^{2+} binding

Introduction

Catalytic RNAs, occurring autonomously or in complex with auxiliary proteins are called ribozymes and can be divided into three classes: (1) the small self-cleaving ribozymes; (2) the large ribozymes, mostly involved in splicing; and (3) the ribosome [1, 2]. Until recently, small ribozymes were known to occur only in the genomes of virus satellites, prokaryotes, and lower eukaryotes. The hairpin, hammerhead, and hepatitis delta virus (HDV) ribozymes play an important role in satellite genome replication, but are found also in transcripts of lower eukaryotes. The *glmS* ribozyme is found in *Bacillus subtilis*, and the Varkud satellite ribozyme is located in mitochondrial transcripts of the mould *Neurospora crassa* [3]. Only recently, small ribozymes were discovered in mammalian genomes, the CoTC motif in the 3′ untranslated region of the β-globin gene [4], a discontinuous hammerhead ribozyme in the 3′ untranslated region of C-type lectin type 2 genes [5], a hammerhead ribozyme in an intron of a tumor suppressor gene [6], and the CPEB3 ribozyme. The latter ribozyme is a self-cleaving, noncoding RNA, highly conserved among mammalian genomes and located in the second intron of the *Cpeb3* gene [7, 8]. This gene encodes a cytoplasmic polyadenylation element binding protein that promotes the elongation of the polyadenine tail of messenger RNA, mediates germ cell development and synaptic plasticity, influences learning and memory, and has been suggested to adopt prion-like conformations [1]. The CPEB3 protein is rather well studied, yet surprisingly little is known about the CPEB3 ribozyme. Most of the available information is based on

M. Skilandat and M. Rowinska-Zyrek contributed equally to this work.

An interactive 3D complement page in Proteopedia is available at http://proteopedia.org/wiki/index.php/Journal:JBIC:24.

Electronic supplementary material The online version of this article (doi:10.1007/s00775-014-1125-6) contains supplementary material, which is available to authorized users.

M. Skilandat · M. Rowinska-Zyrek · R. K. O. Sigel (✉)
Department of Chemistry, University of Zurich,
Winterthurerstrasse 190, 8057 Zurich, Switzerland
e-mail: roland.sigel@chem.uzh.ch

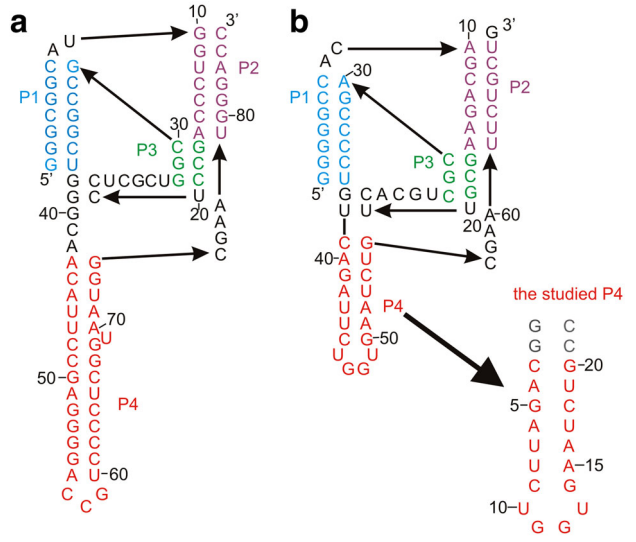

Fig. 1 Secondary structure of the hepatitis delta virus ribozyme (*left*) [8] and the CPEB3 ribozyme (*right*) [7]. The different helical domains are represented by different colors: P1 *blue*, P2 *violet*, P3 *green*, and P4 *red*. The secondary structure of the P4 fragment studied in this work is enlarged, and additional G–C base pairs at the helix end are colored *gray*

comparative studies with the HDV ribozyme. Biochemical analyses have shown strong parallels in catalytic requirements of the HDV and CPEB3 ribozymes [7, 9]. Despite diverging sequences, the HDV and CPEB3 ribozymes can be folded into a highly similar double-pseudoknot structure (Fig. 1), with the active site being located in analogous regions [7, 10].

The HDV ribozyme self-cleaves, yielding a 2′,3′-cyclic phosphate and a 5′-hydroxyl terminus. The suggested mechanism involves an active-site cytosine (C75), located in the J4/2 strand (Fig. 1), which is initially protonated. Its pK_a is shifted more than two pH units toward neutrality [11], and C75N3 can therefore participate in proton transfer reactions at neutral pH. Analogous to C75 in the HDV ribozyme, the CPEB3 ribozyme has an active-site cytosine (C57) in the J4/2 strand, crucial for self-cleavage activity [7]. The catalytic mechanisms of both the HDV ribozyme and the CPEB3 ribozyme strictly require the presence of divalent metal ions [7, 12]. The natural cofactor is Mg^{2+}, which, in the case of the HDV ribozyme, interacts with both the 2′-hydroxyl attacking group and the cleavage-site phosphate to help position the substrate and to activate the nucleophile. Molecular dynamics simulations suggest two types of Mg^{2+} ions associated with the ribozyme near its active site: an inner-sphere bound one and an outer-sphere bound one [13]. The inner-sphere bound ion seems to contribute to catalysis, and the outer-sphere bound ion seems to contribute to stability. In the catalytic mechanism, the protonated C75 acts as a general acid and Mg^{2+} acts as a Lewis acid [14]. The biological meaning of the self-cleavage of the CPEB3 ribozyme is unknown. In humans, a correlation between the rate of self-cleavage, which is tuned by a single-nucleotide polymorphism at position 36, and episodic memory performance has been established [15]. This correlation is attributed to the fact that different cleavage rates would influence the concentrations of the translated CPEB3 protein, which is known to act in memory formation of different model organisms.

The CPEB3 ribozyme consists of four helices (P1–P4; Fig. 1), an extra base pair (P1.1) extending the P1 helix, an internal loop (L3), and two joining regions between helices (J1/2 and J2/4). J2/4 contains the putative active-site cytosine C57. Directly adjacent to the active site is the P4 hairpin (which is the only separate helix), which is not involved in forming the pseudoknot. The sequence of this hairpin is strictly conserved among all CPEB3 sequences of placental mammals investigated so far [7]. To date, the role of this part of the construct remains unknown. Here we focus on this particular region for several reasons: first, detailed knowledge of the P4 structure and metal ion binding sites may provide important insights into its function; second, knowledge of the structure of this well-conserved and structurally separate entity that is in direct proximity to the active site is a first step for understanding the entire structure of the 67-nucleotide-long CPEB3 ribozyme; third, the interactions of this new tetraloop structure with Mg^{2+} are of particular interest from the point of view of bioinorganic chemistry.

Hairpin structures are one of the most important and widespread secondary structure elements found in RNA and are involved in a variety of RNA functions, for example, the mediation of intermolecular interactions with proteins or other nucleic acids or as nucleation sites for RNA folding [7, 16]. Among all hairpin loops, tetraloops are the most widespread ones in RNA [17]. Some well-described examples of tetraloop motifs are GAAA [18, 19] and UUCG [20, 21] (or, more generally, the most widespread motifs are UNCG, GNRA, and CUYG tetraloop sequences, where N represents any nucleotide, R represents a purine, and Y represents a pyrimidine [22, 23]). Determination of the structure of RNA tetraloops provides explanations for their, in some cases, remarkably high stability. To the best of our knowledge, the structure of the UGGU tetraloop has not yet been solved. The most well known UGGU motif is found close to the RNase III cleavage site of a small nuclear RNA of *Saccharomyces cerevisiae* [24–27]. This work now provides the solution structure and resonance assignment of the P4 hairpin closed by a UGGU tetraloop.

Subsequent to the structure determination, we focus on the influence of metal ion binding to the P4 hairpin of the CPEB3 ribozyme, localizing Mg^{2+} and $[Co(NH_3)_6]^{3+}$ binding sites. The detection of Mg^{2+} coordination to

nucleic acids is particularly challenging, since this metal is spectroscopically silent and its binding is kinetically labile [28, 29]. Despite this fact, together with the solution structure, insights into Mg^{2+} binding can be gained by monitoring the effects of Mg^{2+} coordination on the chemical shifts and resonance line widths. Unlike monovalent ions such as Na^+ and K^+, which usually act as bulk electrolytes stabilizing the surface charge between the negatively charged phosphate groups, Mg^{2+} binds with higher affinity, sometimes to very well defined sites, and often plays a specific role in the structure and/or catalysis of complex RNA enzymes [30]. The coordination geometry of Mg^{2+} is strictly octahedral, with a high preference for hard oxygen ligands. Binding of Mg^{2+} can be either direct (inner sphere) to base oxygens, nitrogen, ribose hydroxyl groups, or phosphates, with the backbone phosphate oxygens being the preferred ligands [28, 31], or mediated by coordinated water molecules (outer sphere), which is the commoner way of binding [32]. Site-specific coordination of Mg^{2+} to nucleotide bases, in particular to guanine, is commonly observed [28, 33–35]. $[Co(NH_3)_6]^{3+}$ is a common mimic of $[Mg(H_2O)_6]^{2+}$ and can thus be used to monitor the outer-sphere complexation of Mg^{2+} to nucleic acids [36–38]. With respect to the strong parallels between the HDV ribozyme and the CPEB3 ribozyme, there is reason to consider that the CPEB3 ribozyme is also an obligate metalloribozyme, making its P4 region a fascinating target for studies of Mg^{2+} binding.

Materials and methods

Materials

DNA oligonucleotide templates were purchased from Microsynth (Switzerland), and nucleoside 5′-triphosphates were purchased from GE Healthcare (Switzerland) and Amersham Biosciences (UK). The T7 RNA polymerase used for in vitro transcription was produced in-house according to standard procedures [39]. The Elutrap electroelution apparatus was from Whatman (UK). For desalting, Vivaspin® concentrators (3,000 molecular weight cutoff) from Sartorius-Stedim biotech (France) were used. D_2O (100 %) was purchased from Armar Chemicals (Switzerland).

NMR sample preparation

The P4 RNA was synthesized by in vitro transcription with T7 polymerase from a double-stranded DNA template as previously described [39]. In the template strand, two 2′-*O*-methyl-modified guanine residues were introduced in order to avoid 3′ overhang of the transcribed RNA [40]. Transcription mixtures contained 0.9 μM double-stranded DNA template, 35 mM $MgCl_2$, and each nucleoside 5′-triphosphate at 5 mM concentration. The amount of T7 RNA polymerase was adjusted according to the activity of each enzyme batch. Transcription was allowed to proceed at 310 K for 6–8 h. The transcribed RNA was purified by denaturing 18 % polyacrylamide gel electrophoresis, UV-shadowed, excised from the gel, and recovered by electroelution. By ultrafiltration in Vivaspin® devices, the RNA was washed repeatedly with 1 M KCl, pH 8 to remove tris(hydroxymetghyl)aminomethane and afterwards with water. After lyophilization, the sample was dissolved in 250 μL D_2O or 90 % H_2O/10 % D_2O containing 50 mM KCl and 10 μM EDTA. The pH was adjusted to 6.8 in H_2O or 6.4 in D_2O, corresponding to pD 6.8 [41]. The RNA concentration of the samples ranged between 0.7 and 0.8 mM and was determined using a Varian Cary 100 Scan UV–vis spectrometer by using an extinction coefficient of 281.5 mM^{-1} cm^{-1} at 260 nm. Before acquisition of NMR data, the hairpin was annealed by a 2-min incubation at 358 K, followed by rapid cooling in icy water.

NMR spectroscopy

All spectra were recorded with a Bruker Avance 600-MHz spectrometer with a 5-mm TCI CryoProbe inverse triple-resonance probehead with a *z*-gradient coil, with a Bruker Avance 700-MHz spectrometer with a 5-mm TXI CryoProbe inverse triple-resonance probehead with a *z*-gradient coil, or for 1D ^{31}P spectra with a Bruker Avance 500-MHz spectrometer with a 5-mm QNP CryoProbe probehead. Nonexchangeable proton resonances were assigned from [^1H,^1H] nuclear Overhauser effect (NOE) spectroscopy (NOESY) spectra recorded in 100 % D_2O with a mixing time of 250, 120, or 60 ms at 293, 298, and 303 K. Suppression of the residual water signal was achieved by presaturation pulses. To obtain information on the ribose sugar puckers, [^1H,^1H] total correlation spectroscopy (TOCSY) spectra with 50-ms mixing time were recorded. Exchangeable protons were assigned using [^1H,^1H]-NOESY spectra with a water suppression through gradient-tailored excitation pulse sequence recorded in 90 % H_2O/10 % D_2O at 278 K. 4,4-Dimethyl-4-silapentane-1-sulfonic acid was used as a direct, external reference for ^1H resonances. All spectra were processed with TopSpin 3.0 (Bruker BioSpin, Switzerland), and assignments were performed using the program Sparky (http://www.cgl.ucsf.edu/home/sparky/).

Mg^{2+} and $[Co(NH_3)_6]^{3+}$ titrations

For Mg^{2+} titrations, a 0.72 mM P4 sample in 100 % D_2O was titrated at 298 K with $MgCl_2$ in steps of 0, 0.5, 1, 1.5, 2, 2.5, 3.5, 8, 12, and 18 mM, and a [^1H,^1H]-NOESY spectrum was recorded at each step. All spectra were

assigned using Sparky, and chemical shift changes of the aromatic and sugar protons were analyzed (see also [42]).

A P4 sample in 90 % H_2O/10 % D_2O was titrated with $[Co(NH_3)_6]Cl_3$ in steps of 0, 0.5, 1, 1.5, 2, and 2.5 mM, and a [^{1}H,^{1}H]-NOESY spectrum of the imino proton region at 278 K was recorded at each step. In addition, a [^{1}H,^{1}H]-NOESY spectrum with an excitation sculpting pulse sequence for water suppression was recorded in the presence of 1.5 and 2.5 mM $[Co(NH_3)_6]Cl_3$ at 298 K. Cross peaks between RNA protons and $[Co(NH_3)_6]^{3+}$ protons were assigned in each spectrum. Chemical shift changes caused by $[Co(NH_3)_6]^{3+}$ were determined by titrating a sample with 0, 0.5, 1, 1.5, and 2 mM $[Co(NH_3)_6]^{3+}$ in the same way as described for Mg^{2+} (see above).

Structure calculations and analysis

NOE distances were calculated from the peak volumes in the [^{1}H,^{1}H]-NOESY spectrum of nonexchangeable protons (298 K, 250-ms mixing time) and in the [^{1}H,^{1}H]-NOESY spectrum of exchangeable protons (278 K, 150-ms mixing time). Not overlapping or only moderately overlapping peaks were integrated using Sparky, and distances were calibrated to the fixed H1′–H2′ distance (2.8–3.0 Å) and H5–H6 distance (2.4 Å) of pyrimidines using DYANA's CALIBA macro [43]. According to the result, cross peaks were assigned to four categories: strong (1.8–3.0 Å), intermediate (1.8–4.5 Å), weak (3–6 Å), or very weak (4–7 Å).

Sugar pucker torsion angle restraints were set according to intraresidue H1′–H2′ and H1′–H3′ cross peak intensities in [^{1}H,^{1}H]-TOCSY spectra. Residues with strong cross peaks (G11 and U13) were confined to the south (C2′-*endo*) conformation ($\delta = 145 \pm 20°$, $\nu_1 = 25 \pm 20°$, $\nu_2 = -35 \pm 20°$), and residues with absent cross peaks (G2–C9, G14–C21) were restrained to the north (or C3′-*endo*) conformation ($\delta = 85 \pm 20°$, $\nu_1 = -25 \pm 20°$, $\nu_2 = 37 \pm 20°$). The sugar puckers of G1, U10, C22, and G12 were not restrained as they showed intermediate H1′–H2′ cross peak intensities, or the cross peak pattern did not allow a clear decision to be made (see "Results" and "Discussion"). For the RNA residues in helical regions with C3′-*endo* sugar puckers, the backbone torsion angles α, β, γ, ε, and ζ were set to the values of classic A-form helix ($\alpha = -68°$, $\beta = 178°$, $\gamma = 54°$, $\varepsilon = -153°$, $\zeta = -71°$, all $\pm 20°$). χ angles were set to $-160 \pm 20°$, except for residues G11 and G12, which were set to the *syn* conformation (χ angles set to $60 \pm 20°$), since their intraresidue H1′–H8 cross peaks in a [^{1}H,^{1}H]-NOESY spectrum with 60-ms mixing time were extraordinarily intense. U13 was confined neither to *syn* nor *anti* as the H1′–H6 cross peak was very intense but still less intense than what is usually expected for a *syn* conformation. The α and ζ angles in the loop were left unrestrained as the ^{31}P spectrum of P4 displayed resonances with unusual downfield shifts of 0–1.5 ppm. Base pair formation was validated by the presence of characteristic interstrand [^{1}H,^{1}H]-NOESY cross peaks. In calculations, base pairs were maintained by distance restraints between donor hydrogen and acceptor and between donor and acceptor atoms, and planarity of the base pairs was enforced.

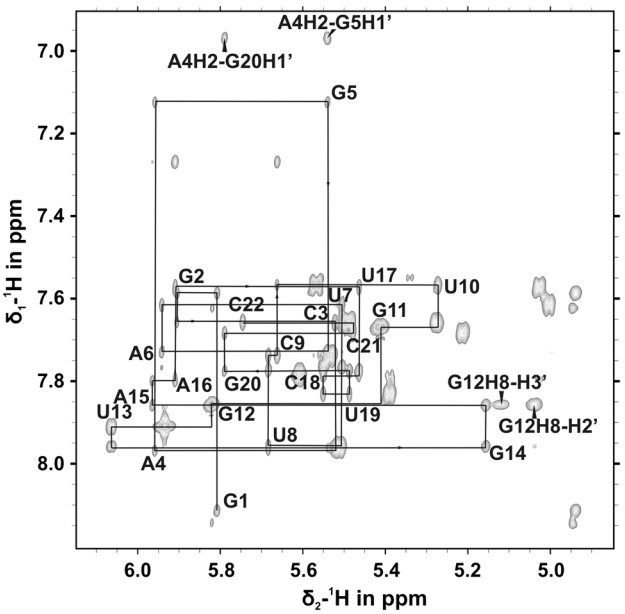

Fig. 2 Sequential walk region in a [^{1}H,^{1}H] nuclear Overhauser effect (NOE) spectroscopy (NOESY) spectrum of the P4 hairpin of the CPEB3 ribozyme (D_2O, 298 K, pD 6.8). Sequential connections between H1′ and H6/8 protons of adjacent bases are traced by *lines*. The resonance of A4H2 appears shifted upfield owing to strong stacking interactions

From the extended RNA sequence, 150 starting structures were calculated by restrained molecular dynamics with CNS version 1.21 [44, 45]. A high-temperature stage of 40 ps at 20,000 K was followed by two cooling stages of 90 ps in torsional space and 30 ps in Cartesian space. The 15 structures of lowest energy were refined using XplorNIH version 2.3 [46, 47] by 88 ps of restrained molecular dynamics cooling from 3,000 to 50 K, calculating 150 output structures. The structure ensemble was analyzed using the programs MOLMOL [48] and PyMOL, and the electrostatic surface potential was determined with the PDBPQR webserver http://nbcr-222.ucsd.edu/pdb2pqr_1.8/) and was visualized using APBSTools2 version 1.4.1 [49] in PyMOL.

Results

Design and NMR spectral features of the P4 domain

The construct used for our NMR studies corresponds to the full-length and wild-type sequence (nucleotides 39–56) of

the conserved mammalian CPEB3 P4 domain. Two G–C base pairs were added to the stem (Fig. 1b) to enhance the stability of the helix, which is rich in A–U base pairs, as well as to provide a suitable starting sequence for in vitro transcription by T7 polymerase. The 22-nucleotide-long hairpin was prepared in high yields by standard in vitro transcription [39].

The [^{1}H,^{1}H]-NOESY spectrum of this P4 construct recorded in D_2O shows well-dispersed resonances in the sequential walk region (Fig. 2). Starting from the 5′ end, all H1′ and H6/H8 protons could be attributed unambiguously by comparing NOESY spectra recorded at different temperatures, and thus showing slightly different chemical shifts.

There are three remarkable features of the [^{1}H,^{1}H]-NOESY spectrum that are discussed in more detail in the following. First, resonances H2 of A4 and H8 of G5 are far upfield at 6.97 and 7.12 ppm, respectively (Fig. 2). These unusually high upfield shifts indicate strong stacking interactions between these bases [18, 50].

The second remarkable feature is the high intensity of the G11H8–H1′ and G12H8–H1′ resonances (Fig. 2), corresponding to the two guanines located in the UGGU tetraloop. Such high intensity of the intranucleotide H1′–H8 cross peak indicates that the nucleotide is in a *syn* conformation, meaning that, unlike in a helix, the nucleobase is rotated around the glycosidic bond and lies above the ribose residue. The χ angle was restrained accordingly in the structure calculation.

Third, strong downfield shifts of the H2′ and H3′ protons of G12 (5.04 and 5.12 ppm, respectively; Fig. 2) are observed. [^{1}H,^{1}H]-TOCSY spectra show very intense G12 H1′–H2′ and H3′–H4′ cross peaks, but only an intermediate H1′–H3′ cross peak. As this rules out both a pure C2′-*endo* and a pure C3′-*endo* conformation [51], no sugar pucker restraints were set for G12. In contrast, the neighboring G11 and U13 H1′–H2′ and H1′–H3′ correlations were very strong, and thus the sugar puckers of these residues were restrained to C2′-*endo*.

Resonances of the imino protons of guanine and uracil are only detectable when they are involved in a hydrogen bond as this considerably slows down the rates of exchange with the solvent. Accordingly, we used imino proton spectra to confirm base pair formation within the helical stem. The [^{1}H,^{1}H]-NOESY spectrum of P4 in 90 % H_2O/10 % D_2O displays strong diagonal peaks for the imino protons of each of the nine Watson–Crick base pairs of the P4 helix as well as the expected cross peaks between neighboring protons on the same strand and on the opposite strand (see also wide infra, Fig. 5b). This confirms a stable formation of the helical stem of P4 under the experimental conditions. Remarkably, G14H1, located in the base pair closing the tetraloop, displays cross peaks to H5, H6, and H1′ of U10, the latter being quite intense (data not shown), which suggests that U10 is pointing inside towards the U9–G14 base pair.

Table 1 NMR restraints and structural statistics for the CPEB3 P4 hairpin structure

NOE-derived distance restraints	449
Per nucleotide	20.4
Intranucleotide ($j - i = 0$)	145
Internucleotide ($k - i = 1$)	247
Long range ($n - i > 1$)	57
Hydrogen-bond restraints	46
Dihedral angle restraints	157
RMSD[a] (Å; for all heavy atoms vs the mean structure)	
Global	1.24 ± 0.48
Stem (1–9, 14–22)	1.07 ± 0.43
Loop (9–14)	0.16 ± 0.06
Loop (10–13)	0.15 ± 0.08

Statistics are given for the 15 lowest-energy structures from 150 calculated structures.

NOE nuclear Overhauser effect, *RMSD* root mean square deviation

To verify that the P4 region adopts the same fold in the absence and presence of the other ribozyme domains, we recorded spectra of the full-length CPEB3 sequence and overlaid them on the P4 spectra (Fig. S1). Although some regions are difficult to compare owing to extensive overlap in the spectrum of the full-length CPEB3, the chemical shifts and relative cross peak intensities belonging to proton resonances of residues 3–20 of the P4 construct match the ones observed in the spectra of the full-length CPEB3. The few minor differences in the chemical shifts can be explained by the presence of the further domains in the full-length CPEB3. Hence, this shows that the P4 region has the same structure in isolation and in the context of the full-length CPEB3 ribozyme.

P4 is a hairpin with an unusual tetraloop structure

The structure of the P4 construct was calculated on the basis of the NOE-derived distance restraints obtained, as well as additional hydrogen bond and dihedral angle restraints. Details are given in Table 1.

With 20.75 NOEs per residue, the loop region is somewhat better defined than the stem with 19 NOEs per residue (cross peaks between loop and stem residues not being counted). The 15 P4 conformers of lowest energy (Fig. 3a) satisfy all distance and dihedral angle restraints within 0.2 Å/5°. All 15 structures in the ensemble adopt a stable and compact fold composed of the A-form helical stem and the very well defined loop region (Fig. 3b). The high convergence of the loop region among the structural ensemble is expressed in the strikingly low root mean square deviation of

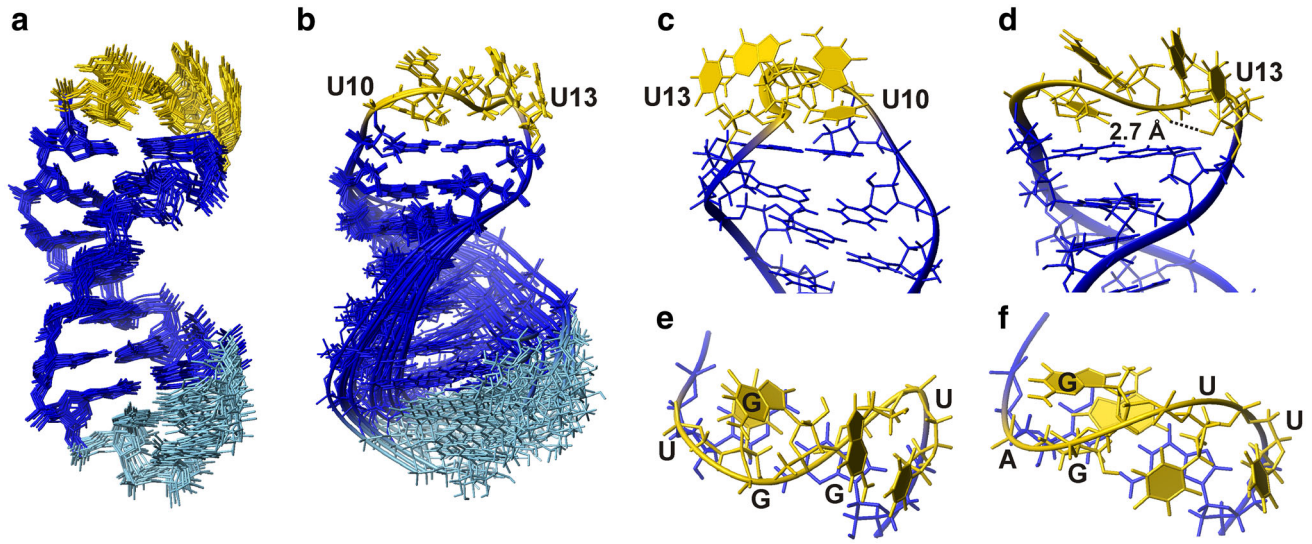

Fig. 3 Solution structure of P4. Fifteen lowest-energy conformers of the CPEB3 P4 region **a** superimposed using all heavy atoms and **b** superimposed using the heavy atoms of residues C9–G14. Owing to the wealth of NOE correlations in the tetraloop (see the main text), this region could be refined to high precision. **c–f** Details of the UGGU tetraloop structure and comparison with the AGUU tetraloop: **c** The loop base U10 stacks onto C9 of the stem helix, whereas G11 and U13 are exposed to the solvent, with their Watson–Crick edges facing different sides. **d** Internal stabilization of the loop is achieved by stacking interactions and a putative hydrogen bond between G12O2P and U13OH2′. **e**, **f** View down the stem helix axis of **e** the UGGU tetraloop and **f** the AGUU tetraloop (prepared from Protein Data Bank entry 1K4B [26]). The *panels* were prepared with MOLMOL [48]

the loop residues compared with the ones of the entire construct and the helical stem alone (Table 1).

The conformation of the loop residues is such that U10, the first residue in the loop, points into the major groove of the stem helix and stacks onto C9. G11, in turn, stacks onto U10, having its Watson–Crick edge exposed to the solvent (Fig. 3c). G12 and U13 also expose their Watson–Crick edges to the solvent, but on the opposite side of the loop pointing towards the minor groove of C9–G14. The characteristic geometry of the loop is largely determined by both G11 and G12 being in a *syn* conformation and having γ angles in the unusual *trans* and *anti* range, respectively. G11 and G12 have an S-type sugar pucker, the conformation of G11 is C2′-*endo*, and the G12 C1′-*exo* sugar conformation is very similar to that of C2′-*endo*. This is typical for tetraloop structures, where the S-type conformation helps to expand the sugar–phosphate backbone and thus to bridge the two strands of the stem with only four nucleotides. In this special case, U13 is also in a C2′-*endo* conformation (see above). The unusually large chemical shift of H2′ of G12 (see above) is probably the result of ring current effects from the bases of G12 and U13. These are stacked on top of each other, and G12H2′ is situated directly next to them. U13 is oriented perpendicular to G14, thus being in a rather exposed position (Fig. 3c, d). Remarkably, the phosphate of U13 is also a point of a sharp directional change of the backbone (Fig. 3c) and sticks out from the molecular surface (Fig. 4b). Apart from the above-mentioned stacking interactions, the loop structure is stabilized by a hydrogen bond between a nonbridging G12 phosphate oxygen and the 2′-hydroxyl group of U13 (Fig. 3d). This hydrogen bond might stabilize the highly unusual backbone trajectory that is not homogeneously curved, like most tetraloops, but instead is indented in the direction of the helix axis (Fig. 3d).

This tetraloop structure is markedly different from that of the main classes of tetraloops that have been investigated so far. The GNRA, UUCG, and CUYG tetraloops all form a non-Watson–Crick base pair between the first and the last nucleotide of the loop, such as the G–U wobble in the UUCG tetraloop and the sheared G–A base pair in the GAAA tetraloop. Such base pair formation can be excluded in the UGGU tetraloop as no NOEs were observed that indicate hydrogen bonding between U10 and U13 or stacking interactions of U13 and G14. Consequently, U10 and U13 are very distant, in perpendicular orientation to each other and on the other side of the sugar–phosphate backbone in the calculated structure. Also, we did not find any evidence for hydrogen bonding between amino protons and phosphate oxygens which helps to stabilize the structures of several tetraloops such as the AGUU, GAAA, and UUCG tetraloops [26, 52, 53].

Specific binding of Mg^{2+} and $[Co(NH_3)_6]^{3+}$ to the helix and loop

Mg^{2+} ions are usually associated with RNA, where they promote folding, as well as contribute to the structure by

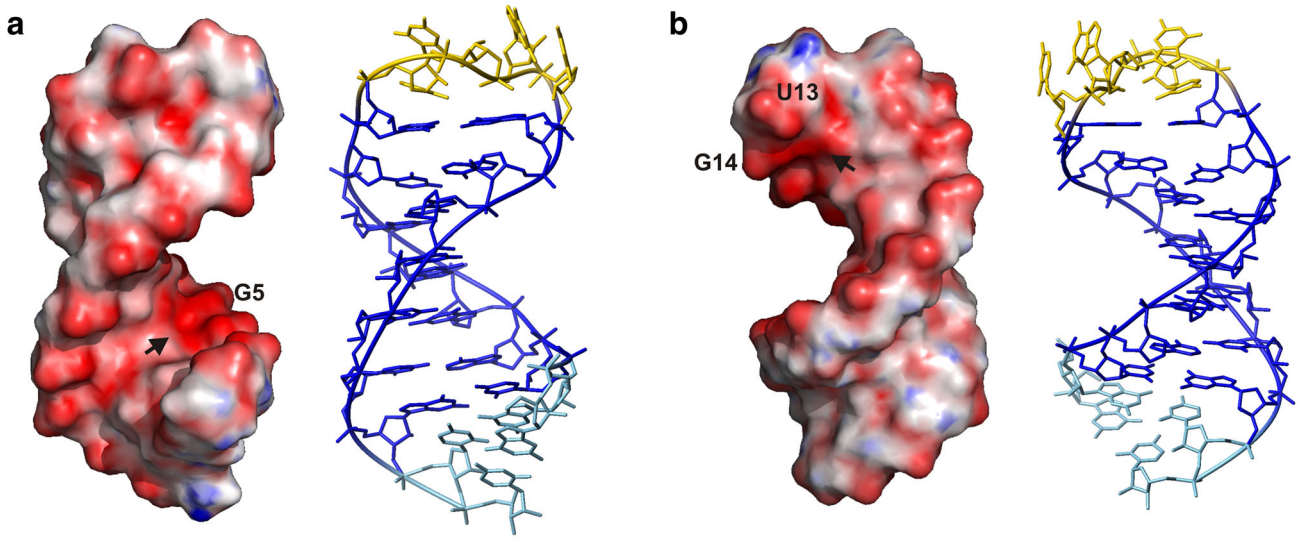

Fig. 4 Electrostatic surface potential map and 3D structure of the lowest-energy conformer seen from opposite sides. The extra G–C base pairs added to the natural sequence are shown in *light blue*, the stem residues are shown in *dark blue*, and the UGGU tetraloop is shown in *yellow*. The electrostatic potential is shown as a gradient from −693 mV (*red*) to 128 mV (*blue*). **a** and **b** are rotated by approximately 180° with respect to each other. The highest density of negative potential is accumulated in two specific sites marked by *arrows*: **a** close to nucleotides A4 and G5, and **b** at the 3′ end of the tetraloop, involving residues G12, U13, and G14. The *panels* were prepared with PyMOL and MOLMOL [48]

stabilizing the 3D structure or mediating interactions with other molecules and/or catalysis [30, 35, 54]. The identification of Mg^{2+} binding sites in RNA molecules is thus crucial to understand RNA function.

We titrated the P4 hairpin with increasing amounts of Mg^{2+} and followed the chemical shift changes of sugar H1′ and nucleobase H2, H6, and H8 by [^{1}H,^{1}H]-NOESY spectroscopy. Such chemical shift changes ($\Delta\delta$) result either from the binding of Mg^{2+} close to the proton, whose resonance is shifted, or from a local Mg^{2+}-induced structural change [29, 55]. In the P4 hairpin, Mg^{2+} binding affects both the chemical shifts and the resonance line widths of the nucleic acid protons. Chemical shift mapping analysis shows that the resonances with the largest chemical shift change are those of (1) the first two nucleotides G1 and G2, (2) A4 and G5 in the middle of the hairpin stem, and (3) G12 and G14 near the 3′ end of the tetraloop. The chemical shift changes on addition of 5 mM Mg^{2+} are shown in Fig. 5a. G1 and G2 H8 and H1′ resonances are broadened to the baseline at 5 mM Mg^{2+}. Strong binding of Mg^{2+} at the 5′-terminal triphosphate with its high negative charge is expected and has been observed in a variety of other studies [56–58].

More interesting is the binding of Mg^{2+} to nucleotides A4 and G5, as indicated by the strong chemical shift changes observed on addition of Mg^{2+} (Fig. 5a) and confirmed by the accumulation of negative charge in the major groove at N7 and the nonbridging phosphate oxygens of A4 and G5 as well as G5O6 (Fig. 4a). Mg^{2+} binding to this part of the stem also causes the strong chemical shift change of U19H6, as U19 forms a base pair with A4. This, however, seems to be an indirect effect of Mg^{2+} binding to G5 and A4 because neither U19 nor neighboring nucleotides form a cluster of negative charge.

Also, the third Mg^{2+} binding site at U13 and G14 shows a high negative surface potential (Fig. 4b). On addition of Mg^{2+}, resonances from nucleotides G12 and G14 are shifted significantly as are those from G11, U13, and A15, but to a lesser extent (Fig. 5a). Accordingly, G11 and A15 marginally contribute to the negative charge cluster that likely attracts the metal ion, but as they are in the proximity of the G12/G14 binding site, they might be exposed to a different chemical environment in the absence and presence of Mg^{2+}.

To better pinpoint the metal ion binding site, we performed [^{1}H,^{1}H]-NOESY experiments with the P4 hairpin in the presence of 1.5 mM $[Co(NH_3)_6]^{3+}$. This stable, exchange-inert complex mimics the fully hydrated (hexaaqua) Mg^{2+} ion and is a commonly used mimic for outer-sphere coordination of Mg^{2+} [37, 59]. All protons of the ammine ligands resonate at one common frequency, and NOE cross peaks can be detected between this resonance and different resonances of RNA protons in the $[Co(NH_3)_6]^{3+}$ binding site (Fig. 5b).

The $[Co(NH_3)_6]^{3+}$ titrations perfectly support the results from the Mg^{2+} titration described above for the two binding sites in the helical stem. Addition of $[Co(NH_3)_6]^{3+}$ leads to a broadening (G1H1), a change in chemical shift (G1H1′, C22H1′), and cross peaks to RNA protons [G2H1 (Fig. 5b), G1H8, G2H8, C21H5, and G20H1′ (data not shown)].

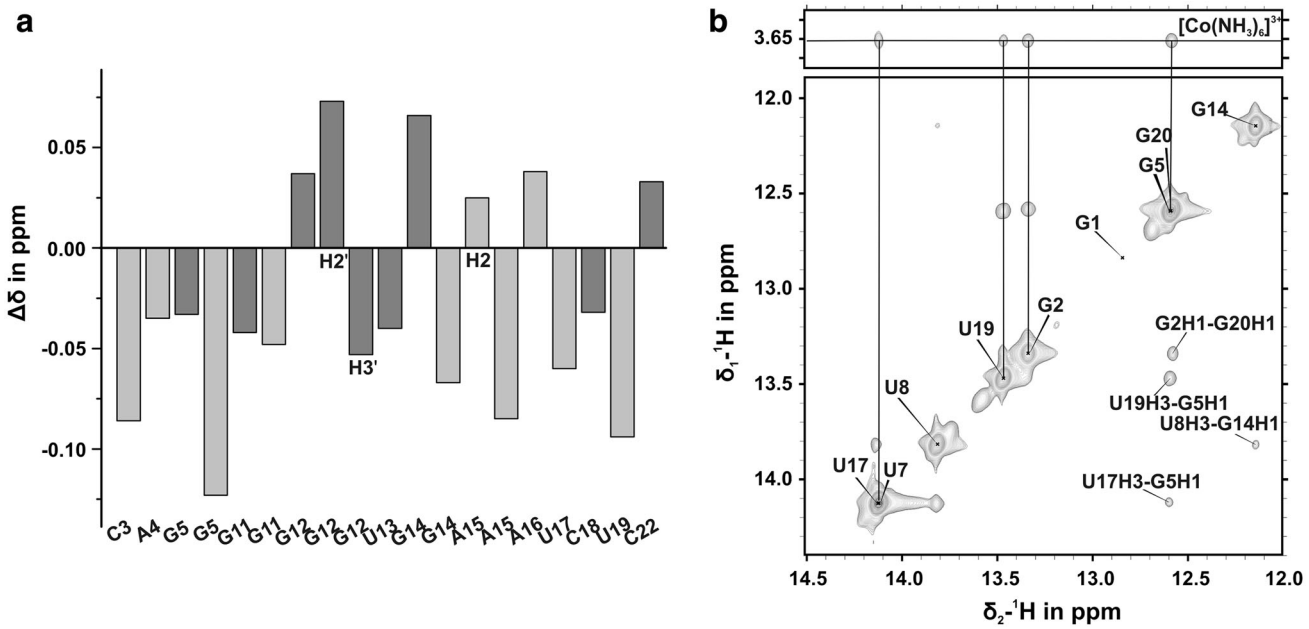

Fig. 5 Localization of Mg^{2+} binding sites in P4. **a** Chemical shift changes $\Delta\delta$ on addition of 5 mM Mg^{2+} compared with the chemical shift in the absence of Mg^{2+} ($\Delta\delta = \delta_{Mg2+} - \delta$). Only residues with $\Delta\delta > 0.02$ ppm are shown. Unless otherwise labeled, *light-gray bars* represent $\Delta\delta$ of H6/8 protons and *dark-gray bars* represent $\Delta\delta$ of H1′ protons. **b** [^{1}H,^{1}H]-NOESY spectrum of the exchangeable protons of P4 in the presence of 1.5 mM $[Co(NH_3)_6]^{3+}$ recorded in 90 % H_2O/10 % D_2O at 278 K

Aside from Mg^{2+} binding to the 5′-triphosphate, specific outer-sphere coordination of Mg^{2+}/$[Co(NH_3)_6]^{3+}$ also occurs within the stem. Distinct cross peaks between the protons of $[Co(NH_3)_6]^{3+}$ and G5H1, U17H3, and U19H3, all located in the stem helix on both strands, are detected (Fig. 5b). The [^{1}H,^{1}H]-NOESY spectrum of the nonexchangeable proton region further confirms the existence of a specific Mg^{2+} binding site in the proximity of G5, corroborated by correlations with G5H8, A4H8, and C18H41 (in a base pair with G5) (data not shown).

The combined titration experiments described above support a specific outer-sphere coordination of Mg^{2+} in the central region of the P4 stem, which is known to occur in RNA A-form helices [18, 21]. In contrast, $[Co(NH_3)_6]^{3+}$ affects neither any exchangeable nor any nonexchangeable resonances of the UGGU tetraloop and the closing base pair. Consequently, Mg^{2+} binds most likely in a partly inner-sphere manner to the loop region, which cannot be mimicked by $[Co(NH_3)_6]^{3+}$. The sharp turn in the sugar–phosphate backbone at U13 causes all suitable ligands for Mg^{2+}, which are the phosphates of G12, U13, and G14, as well as G14N7 or G14O6, to be situated close to each other, too close, in fact, to accommodate the large $[Co(NH_3)_6]^{3+}$ complex.

Discussion

By NMR spectroscopy, we solved the solution structure of a well-defined and conserved fragment of the human CPEB3 ribozyme, the P4 hairpin, and for the first time characterized the structure of a UGGU tetraloop in detail. UGGU tetraloops have been found to appear in targets of the *Saccharomyces cerevisiae* endonuclease Rnt1p [25]. Rnt1p recognizes many of its targets by interactions with AGNN tetraloops [24], but UGNN tetraloops are also recognized [25, 26]. The most similar tetraloop sequence, for which a 3D structure is available, and which is found in an Rnt1p target, is the AGUU tetraloop [26]. Comparing the structures of the UGGU tetraloop and the AGUU tetraloop (Protein Data Bank ID 1K4B), we find that both loops have an S-shaped backbone trajectory (Fig. 3e, f) as seen from the top, with the first two and the last two loop bases pointing to opposite sides of the helix. Also, the second loop residue (G11 in P4) being the conserved guanine is in a *syn* conformation and has a similar position and orientation in both loops, which is an important determinant for recognition by Rnt1p [24, 26, 60], and is also found in the AGUC and AGAA tetraloops that serve as recognition sites for Rnt1p [60]. The crystal structure of a 32mer hairpin containing the AGAA tetraloop bound to the RNA-binding domain of Rnt1p revealed that the *syn* guanine is important for the overall fold of the tetraloop which is recognized by Rnt1p [61] rather than the *syn* guanine itself.

Of all available tetraloop sequences, the UGUU tetraloop [62] is the most similar one. This tetraloop also plays a role in RNA–protein interaction; it is found in hepatitis beta virus encapsidation signals and mediates the initiation

of replication when it is bound by the viral reverse transcriptase [63]. Despite the large sequence similarity of the two tetraloops and their closing base pair (cUGGUg vs cUGUUg), the tetraloop fold is remarkably different. The guanine in UGUU is in an *anti* conformation and the first and last uracils form a buckled *cis*-wobble base pair. The only common features are the stacking of the first uracil on the adjacent cytosine in the hairpin stem and the distribution of the first two bases pointing towards the major groove of the stem and the third loop base pointing towards the minor groove. It is quite striking how much structural variation is introduced by the substitution of a single nucleobase. This underlines the variability of RNA folding even in such small motifs as a tetraloop.

Another interesting aspect of the UGGU loop structure is the fact that it forms an inner-sphere binding site for Mg^{2+} that probably stabilizes this particular fold of the tetraloop. This feature, the similarity to the AGNN tetraloop, the sequence conservation of P4 between mammalian transcripts, and the fact that it is the only protuberant domain of the otherwise very compact ribozyme structure point out that P4 has some functional relevance, for example, as an interaction site with a protein or RNA binding partner. However, to date there are no biological or biochemical data on the function of this region that could support this hypothesis.

Apart from the tetraloop itself, the P4 stem also deserves some attention, as it is very close to the active-site cytosine C57 (Fig. 1b). With regard to this, it is possible that the metal ion coordination at A4 and G5 might be relevant for the active-site structure. To further investigate this, we analyzed different crystal structures of the HDV ribozyme that have the same general secondary structure as the P4 region directly preceeding the catalytic cytosine. In the structures of the precleavage HDV ribozyme [14] and the C75U mutant [64] there is a major-groove binding site, accessible to both Mg^{2+} and $[Co(NH_3)_6]^{3+}$ in the upper part of the P4 stem. The respective Mg^{2+} ion is rather far from C75, but still close enough to interact with the phosphates of C41 and A42 of the P1.1 region and thus might support the continuous stacking interactions between P1, P1.1, and P4, which are present in the HDV ribozyme. A similar function of the P4 stem Mg^{2+} binding site in the CPEB3 ribozyme seems reasonable to suggest.

Acknowledgments Financial support from a Sciex postdoctoral grant (no. 11.156, to M.R.-Z.), a Marie Curie fellowship (no. PIEF-GA-2012-329700 to M.R.-Z.), the Swiss National Science Foundation (to R.K.O.S.), and the University of Zurich is gratefully acknowledged. R.K.O.S. is a recipient of a European Research Council Starting Grant (microRNA). Structure coordinates have been deposited in the RCSB Protein Data Bank (ID 2M5U), and chemical shifts have been deposited in the Biological Magnetic Resonance Bank with accession code 19081.

References

1. Luptak A, Szostak JW (2007) In: Lilley DMJ, Eckstein F (eds) Mammalian self-cleaving ribozymes. Royal Society of Chemistry, Cambridge, pp 123–133
2. Sigel A, Sigel H, Sigel RKO (eds) (2011) Structural and catalytic roles of metal ions in RNA. Metal ions in life sciences, vol 9. Royal Society of Chemistry, Cambridge
3. Hammann C, Hartmann RK, Marchfelder A (2007) Biol Chem 388:659–660
4. Teixeira A, Tahiri-Alaoui A, West S, Thomas B, Ramadass A, Martianov I, Dye M, James W, Proudfoot NJ, Akoulitchev A (2004) Nature 432:526–530
5. Martick M, Horan LH, Noller HF, Scott WG (2008) Nature 454:899–902
6. de la Pena M, Garcia-Robles I (2010) EMBO Rep 11:711–716
7. Salehi-Ashtiani K, Luptak A, Litovchick A, Szostak JW (2006) Science 313:1788–1792
8. Talini G, Branciamore S, Gallori E (2011) Biochimie 93:1998–2005
9. Chadalavada DM, Gratton EA, Bevilacqua PC (2010) Biochemistry 49:5321–5330
10. Webb CHT, Luptak A (2011) RNA Biol 8:719–727
11. Gong B, Chen JH, Chase E, Chadalavada DM, Yajima R, Golden BL, Bevilacqua PC, Carey PR (2007) J Am Chem Soc 129: 13335–13342
12. Nakano S-I, Proctor DJ, Bevilacqua PC (2001) Biochemistry 40:12022–12038
13. Veeraraghavan N, Ganguly A, Golden BL, Bevilacqua PC, Hammes-Schiffer S (2011) J Phys Chem B 115:8346–8357
14. Chen J-H, Yajima R, Chadalavada DM, Chase E, Bevilacqua PC, Golden BL (2010) Biochemistry 49:6508–6518
15. Vogler C, Spalek K, Aerni A, Demougin P, Muller A, Huynh KD, Papassotiropoulos A, de Quervain DJ (2009) Front Behav Neurosci 3:4
16. Cheong C, Cheong H-K (2010) In: Encyclopedia of life sciences. Wiley, Chichester. doi:10.1002/9780470015902.a0003135.pub2
17. Deng NJ, Cieplak P (2010) Biophys J 98:627–636
18. Korth MMT, Sigel RKO (2012) Chem Biodiv 9:2035–2049
19. Sigel RKO, Sashital DG, Abramovitz DL, Palmer AG, Butcher SE, Pyle AM (2004) Nat Struct Biol 11:187–192
20. Nozinovic S, Furtig B, Jonker HRA, Richter C, Schwalbe H (2010) Nucleic Acids Res 38:683–694
21. Donghi D, Pechlaner M, Finazzo C, Knobloch B, Sigel RKO (2013) Nucleic Acids Res 41:2489–2504
22. Fürtig B, Richter C, Bermel W, Schwalbe H (2004) J Biomol NMR 28:69–79
23. Banas P, Hollas D, Zgarbova M, Jurecka P, Orozco M, Cheatham TE, Sponer J, Otyepka M (2010) J Chem Theor Comput 6:3836–3849
24. Chanfreau C, Buckle M, Jacquier A (2000) Proc Natl Acad Sci USA 97:3142–3147
25. Seipelt RL, Zheng BH, Asuru A, Rymond BC (1999) Nucleic Acids Res 27:587–595
26. Wu HH, Yang PK, Butcher SE, Kang S, Chanfreau G, Feigon J (2001) EMBO J 20:7240–7249
27. Ghazal G, Ge DL, Gervais-Bird J, Gagnon J, Abou Elela S (2005) Mol Cell Biol 25:2981–2994
28. Freisinger E, Sigel RKO (2007) Coord Chem Rev 251: 1834–1851
29. Erat MC, Sigel RKO (2011) Met Ions Life Sci 9:37–100
30. Sigel RKO, Pyle AM (2007) Chem Rev 107:97–113
31. Schnabl J, Sigel RKO (2010) Curr Opin Chem Biol 14:269–275
32. Sreedhara A, Cowan JA (2002) Biometals 15:211–223
33. Ennifar E, Walter P, Dumas P (2003) Nucleic Acids Res 31:2671–2682

34. Petrov AS, Pack GR, Lamm G (2004) J Phys Chem B 108:6072–6081
35. Sigel RKO, Sigel H (2013) In: Pecoraro VL (ed) Bioinorganic fundamentals and applications: metals in natural living systems and metals in toxicology and medicine. Comprehensive inorganic chemistry II, vol 3. Elsevier, Amsterdam, pp 623–660
36. Cowan JA (1993) J Inorg Biochem 49:171–175
37. Rowinska-Zyrek M, Skilandat M, Sigel RKO (2013) Z Anorg Allg Chem 639:1313–1320
38. Schmitz M (2004) Nucleic Acids Res 32:6358–6366
39. Gallo S, Furler M, Sigel RKO (2005) Chimia 59:812–816
40. Kao C, Rüdisser S, Zheng M (2001) Methods 23:201–205
41. Glasoe PK, Long FA (1960) J Phys Chem 64:188–190
42. Donghi D, Johannsen S, Sigel RKO, Freisinger E (2012) Chimia 66:791–797
43. Güntert P, Mumenthaler C, Wüthrich K (1997) J Mol Biol 273:283–298
44. Brünger AT (1992) X-PLOR. Version 3.1. A system for X-ray crystallography and NMR. Yale University Press, New Haven
45. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Acta Crystallogr D 54:905–921
46. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) J Magn Reson 160:65–73
47. Schwieters CD, Kuszewski JJ, Clore GM (2006) Prog Nucl Magn Reson Spectrosc 48:47–62
48. Koradi R, Billeter M, Wüthrich K (1996) J Mol Graph Model 14:29–32, 51–55
49. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Proc Natl Acad Sci USA 98:10037–10041
50. Sigel H (2004) Pure Appl Chem 76:1869–1886
51. Varani G, Aboulela F, Allain FHT (1996) Prog Nucl Magn Reson Spectrosc 29:51–127
52. Allain FHT, Varani G (1995) J Mol Biol 250:333–353
53. Jucker FM, Heus HA, Yip PF, Moors EHM, Pardi A (1996) J Mol Biol 264:968–980
54. Woodson SA (2005) Curr Opin Chem Biol 9:104–109
55. Pechlaner M, Sigel RKO (2012) Met Ions Life Sci 10:1–42
56. Sigel H, Bianchi EM, Corfù NA, Kinjo Y, Tribolet R, Martin RB (2001) Chem Eur J 7:3729–3737
57. Sigel H, Griesser R (2005) Chem Soc Rev 34:875–900
58. Erat MC, Sigel RKO (2007) Inorg Chem 46:11224–11234
59. Rüdisser S, Tinoco I (2000) J Mol Biol 295:1211–1223
60. Lebars I, Lamontagne B, Yoshizawa S, Elela SA, Fourmy D (2001) EMBO J 20:7250–7258
61. Wu H, Henras A, Chanfreau G, Feigon J (2004) Proc Natl Acad Sci USA 101:8307–8312
62. Girard FC, Ottink OM, Ampt KAM, Tessari M, Wijmenga SS (2007) Nucleic Acids Res 35:2800–2811
63. Nassal M, Schaller H (1996) J Viral Hepatitis 3:217–226
64. Ke A, Zhou K, Ding F, Cate JHD, Doudna JA (2004) Nature 429:201–205